

## **Structured decision-making drives guidelines panels' recommendations 'for' but not 'against' health interventions**

Benjamin Djulbegovic<sup>1,2</sup>, Tea Reljic<sup>3</sup>, Shira Elqayam<sup>4</sup>, Adam Cuker<sup>5</sup>, Iztok Hozo<sup>6</sup>, Qi Zhou<sup>7</sup>, Shelly-Anne Li<sup>8</sup>, Paul Alexander<sup>7</sup>, Robby Nieuwlaat<sup>7</sup>, Wojtek Wiercioch<sup>7</sup>, Holger Schünemann<sup>7</sup> and Gordon Guyatt<sup>7</sup>

<sup>1</sup> Department of Supportive Care Medicine, City of Hope, 1500 East Duarte Rd, Duarte, CA

<sup>2</sup> Department of Hematology, City of Hope, 1500 East Duarte Rd, Duarte, CA;

<sup>3</sup> University of South Florida, 12901 Bruce B Downs Blvd, Tampa, FL;

<sup>4</sup> De Montfort University, Leicester, UK.

<sup>5</sup> Perelman School of Medicine, University of Pennsylvania, Philadelphia, PA

<sup>6</sup> Department of Mathematics, Indiana University, Gary, IN

<sup>7</sup> Department of Health Research Methods, Evidence, and Impact, McMaster University, Hamilton, ON, Canada

<sup>8</sup> Faculty of Nursing, University of Toronto, Canada

Abstract: 200

Word count: 2,955

Correspondence: Benjamin Djulbegovic, MD, PhD

1500 E Duarte Road

Duarte, CA 91010-3000

Phone: 626-218-7502

e-mail: [bdjulbegovic@coh.org](mailto:bdjulbegovic@coh.org)

## Abstract

**Background:** The determinants of guideline panels' recommendations remain uncertain.

**Objective:** To investigate factors considered by members of 8 panels convened by the American Society of Hematology (ASH) to develop guidelines using GRADE system.

**Study Design and Setting:** web-based survey of the participants in the ASH guidelines panels. **Analysis:** two level hierarchical, random-effect, multivariable regression analysis to explore the relation between GRADE and non-GRADE factors and strength of recommendations (SOR).

**Results:** In the primary analysis, certainty in evidence [OR=1.83; (95CI% 1.45 to 2.31)], balance of benefits and harms [OR=1.49 (95CI% 1.30 to 1.69)] and variability in patients' values and preferences [OR=1.47 (95CI% 1.15 to 1.88)] proved the strongest predictors of SOR. In a secondary analysis, certainty of evidence was associated with a strong recommendation [OR=3.60 (95% CI 2.16 to 6.00)] when panel members recommended “**for**” interventions but not when they made recommendations “**against**” [OR=0.98 (95%CI: 0.57 to 1.8)] consistent with “yes” bias. Agreement between individual members and the group in rating SOR varied (kappa ranged from -0.01 to 0.64).

**Conclusion:** GRADE's conceptual framework proved, in general, highly associated with SOR. Failure of certainty of evidence to be associated with SOR against an intervention, suggest the need for improvements in the process.

**Running title:** guidelines panels' decision-making

**Keywords:** Practice Guidelines, Clinical Recommendations, “Yes” bias, Decision Theory, Group Decision Making, GRADE

## What is new?

### Key findings:

- The GRADE guidelines system specifies factors that guidelines panels **should** take into considerations when issuing recommendations. However, many other (non-GRADE) factors may also affect recommendations.
- To what extent GRADE vs. non-GRADE factors influence guidelines panels' decision-making remains uncertain.
- We found that GRADE factors affect guidelines decision-making process more than non-GRADE factors, likely due to the effect of instructions provided within structured GRADE Evidence-to-Decision (EtD) framework. Consistent with principles of evidence-based medicine, we confirmed relation between the certainty of evidence and strength of recommendations.
- The findings remained robust when panels issued recommendations **for** health interventions. However, when the panels generated recommendations **against** health interventions, the relation between certainty of evidence and strength of recommendations disappeared **pointing to the existence of so called “yes” bias (people acquiesce to “yes” statements more readily than to “no” statements)**
- Even within highly structured GRADE process, the panel members demonstrated **variability** in their **individual** responses (kappa between individual panel members and the group consensus vote for strength of recommendations ranged from very poor (-0.01) to moderate (0.64)),
- Depending on the analytical model, some non-GRADE factors were also associated with the strength of recommendations issued by the panels. Different non-GRADE factors were associated with recommendations “for” vs. “against” health interventions. However, age/clinical experience of the panelists remained statistically significant across all models.

### What this adds to what is known:

- This quantitative analysis of 8 panels confirms that GRADE instruction given within EtD structured framework results in consideration of GRADE factors as intended by the GRADE system.
- The system does not, however, appear to give consistent results when the panels issue recommendation **for** vs **against** health intervention.
- In addition, individual member “assessment” often considerably differ from the group, consensus vote.

### What is the implication, and what should change now:

- Guideline panels that place a high value on adherence to the GRADE system should consider use of EtD framework in developing their recommendations.
- **To avoid “yes” bias**, guidelines developers should, in most instances, express all recommendations as a vote “for” instead of “against” recommendations
- Exploration of reasons why panel members are sometimes in agreement and sometimes not may inform the need for additional strategies such as more extensive training in GRADE to reduce variability.

Trustworthy evidence-based clinical practice guidelines (CPG)<sup>1-3</sup> represent one approach to addressing suboptimal clinical decision making.<sup>4,5 6 7</sup> In fact, measuring adherence to CPGs is one of the key approaches to quality improvement.<sup>7 8</sup>

If CPGs are to improve health outcomes, they must be developed using rigorous methodological principles<sup>2</sup> developed during the last 20 years through systems of rating the certainty of evidence and strength of recommendations.<sup>9 10 11 12 13</sup> Of these systems, the GRADE (Grading of Recommendations Assessment, Development, and Evaluation) approach represents the most transparent, rigorously developed and documented to date,<sup>14 2</sup> which is endorsed by over 100 professional organizations, including the World Health Organization, the Cochrane Collaboration, and a number of leading American organizations.<sup>15</sup>

GRADE has identified a number of factors that CPG panels *should* consider when making recommendations, including the certainty of evidence, the balance between benefits and harms, patient values and preferences, resource and cost considerations, as well as issues related to acceptability, feasibility, and health equity.<sup>16,17</sup> Although GRADE provides a normative system for how CPG panels 'ought to' develop guidelines, in what manner guideline panels *actually* make their judgments remains unclear.

Despite the breadth of GRADE's specified considerations, many additional factors not formally captured in the GRADE system may affect panel judgments. Broadly, these factors include<sup>18</sup>: a) *decision features, or characteristics of the decision/recommendation* (e.g., high stake vs. low-stake clinical recommendations, such as developing guidelines for vulnerable populations in a politically charged atmosphere), b) *situational/contextual factors* (e.g., time pressure, cognitive load, role

in the panel as chair, methodologist, panelists, etc), and c) *individual characteristics of the decision maker (e.g. age)*.<sup>18-21</sup>

How and to what extent these additional factors contribute to the decision-making process remains unclear. In addition, development of CPGs ultimately relies on the *group judgment* of the panel. Despite its importance, we know little about how the group consensus relates to the individual judgments of its members. We, therefore, designed a study addressing the interplay of group and individual processes in real life decision making and provide the first analysis of a guidelines panel's decision process.

## **Methods**

We studied the process in 8 panels convened by the American Society of Hematology (ASH) to develop guidelines for the management of the following conditions: heparin-induced thrombocytopenia (HIT); thrombophilia; venous thromboembolism (VTE) in pregnancy; VTE in pediatric populations; optimal management of anticoagulation therapy; VTE in patients with cancer; treatment of VTE; and management of immune thrombocytopenia (ITP).

A series of webinars introduced panel members to the GRADE system. During a number of conference calls, the panel members defined and prioritized the clinical questions, guided a systematic review team in the collection and analysis of the relevant evidence, and in some cases discussed pre-voting results for guideline questions. All recommendations included judgments (weak or conditional vs strong) in favor of intervention (I) or comparator (C) for a given outcome (O) related to the population of interest (P).

Each panel developed final recommendations through *group consensus* during face-to-face meetings using GRADE's structured evidence-to-decision (EtD) framework.<sup>16,17</sup> For each recommendation, the panel made explicit judgments for each factor in the framework; before doing so, panelists reviewed a summary of these judgments (see Appendix for the actual presentation framework). Each panel member completed a survey detailing their judgments related to relevant GRADE factors and the final recommendations during the meeting or shortly thereafter.

We used frequencies and percentages to describe characteristics of panels and panel members participating in the development of the ASH guidelines. We explored GRADE and non-GRADE factors that might influence the panels' recommendations, all defined *a priori* as per current literature.<sup>18-21</sup> The GRADE factors included certainty of evidence supporting recommended intervention, balance between benefits and harms, assessment of variability or uncertainty in patients' value and preferences, and resources that the panel judged may be needed to implement recommendations. Certainty of evidence was coded on a 1 to 4 scale, with 1 indicating very low certainty of evidence and 4 high certainty of evidence. Variability or uncertainty in patient's values and preferences (V&P) was coded on a scale 1 to 4 (1=important, 2=possibly important, 3=probably not important, 4=not important). Judgments on use of resource/costs were coded on a 5-point Likert scale (1=large costs, 2=moderate costs, 3=neither, 4=moderate savings, 5=large savings). Judgments on the balance of intervention benefit/harms was coded on 5-point Likert scale (1=favors the comparison, 2=probably favors the comparison, 3=does not favor either, 4=probably favors the intervention; 5=favors the intervention). Each of these categorical variables was treated in the

analysis as continuous assuming the equivalent interval effects among the consecutive scores.

Although the panels, using the EtD framework, also considered issues of acceptability, feasibility, and health equity (see Appendix), limiting response burden on panelists precluded our considering these issues.

Non-GRADE variables included:

a) *individual characteristics of the decision maker*: age, sex, experience, expertise and cognitive styles, i.e. propensities to favor one decision-making or reasoning approach over another.<sup>20</sup> The latter was assessed by administration of instruments to measure objectivism, i.e., tendency to seek empirical information to support decision making; intolerance of uncertainty<sup>22</sup>; maximizing-satisficing, i.e., assessment of tendency for individual to employ reasoning processes that will lead to making a good vs. best possible decision<sup>23</sup>; propensity to engage in analytical, rational thinking vs experiential-intuitive thinking<sup>20,24</sup>; and tendency to experience regret about making a decision.<sup>21</sup> These instruments have proved valid and applicable to assessment of physicians' decision-making.<sup>20</sup>

b) *characteristics of the decision/recommendation*: recommendations made for vulnerable populations (children, women, inner city, rural, ethnic minority, low-income), reports of feeling pressured to issue certain type of recommendations/to conform with the group due to the potentially politically sensitive nature of guideline recommendations.

c) *situational/contextual factors related to a given guideline recommendation*: individual panel member's conflicts of interest, role in the panel (chair, methodologist,

patient representative, panel member). The Supplementary material provides details related to all variables and instruments.

We constructed a model relating these variables to the strength of recommendations as either strong, weak, or no recommendation. We repeated the analysis according to the direction of the recommendation (“for” vs. “against”), omitting questions in which panels did not make a recommendation.<sup>25,26</sup>

We employed a two level hierarchical, mixed multivariable logistic and ordered regression analysis to account both for panel level factors and individual level factors. Thus, judgments of recommendations were clustered within panel members, and these were clustered within panels.

To compare individual panel recommendations with the group consensus, we calculated the agreement (*kappa* statistics and correlations) between each individual panel member’s average judgment (weighted by the number of their responses) and the group consensus recommendation. To account both for sampling error and the variability among the panels, we pooled *kappa* statistics across all panels by meta-analyzing it under a random-effects model.<sup>27</sup> To estimate the random effects of the panels on the percentage of the total residual variance in each individual member’s voting pattern, we estimated intraclass correlations (ICCs) after running the two-level mixed effect logistic regressions. All calculations were performed using STATA, version 15<sup>28</sup>, and verified in SAS, version 9.4, by a second investigator.

## **Results**

Fig 1 presents an overview of the data collection process. Table 1 presents characteristics of the panels and panel members participating in the development of the



ASH guidelines. Typical panelists were male hematologists around 50 years of age from the United States with approximately 20 years of clinical experience.

The panel meetings occurred between November 2016 and August 2017 in Washington, D.C. and lasted between 15 and 26 hours across two days (median=10 hours per day).

Of 21 variables potentially associated with the strength of recommendations, 3 GRADE and 2 non-GRADE factors displayed statistically significant association at the conventional  $p < 0.05$  levels (Table 2). Panel members' judgment of certainty of the evidence [OR=1.84 (95%CI 1.46 to 2.31)] proved the strongest predictor - the more confident the panel members were regarding the certainty of the evidence, the more inclined they were to issue strong recommendations.

Other factors associated with strong recommendations included age (per decade) [OR=1.79 (95CI% 1.2 to 2.84)] (older panel members were more inclined to make strong recommendations), followed by balance of benefits and harms [OR=1.49 (95CI% 1.30 to 1.69)] (when balance favors intervention, the panelists are more likely to issue a strong recommendation), the uncertainty or variability in patients' V&P [OR=1.47 (95CI% 1.15 to 1.88)] (the less uncertainty or variability, more likely the panel was to issue a strong recommendation), and intolerance of uncertainties [OR=0.57 (95CI% 0.37 to 0.86)] (more intolerance, less likely a strong recommendations).

Table 3 showed the logistic regression analysis when the panel members issued recommendations **“strong for” vs. “weak for”** in favor of a given intervention. In this analysis, judgement about balance between benefits and harms was associated with an

OR of 18.3 [95% CI 7.67 to 43.7] for recommendations in favor of the intervention. The second strongest predictor was certainty of evidence [OR=3.61 (95%CI 2.17 to 6.01)]. When panels judged that certainty in evidence is high and benefits outweigh harms in favor of intervention over the comparator, the predicted probability of issuing strong recommendation in favor of the intervention exceeded 90% (Fig 2).

Assessment of patients' values and preferences as well as consideration of costs/resources were also highly statistically significant but at a somewhat lower odds ratio (Table 3). Methodologists, in comparison to panel chairs, were less likely to issue strong recommendations [OR=0.06 (95CI% 0.04 to 0.85)]. Three non-GRADE factors also show statistically significant or borderline significant associations (Table 3). As in the main analysis (Table 2), older panel members were more inclined to make strong recommendations [OR=2.6 (95% CI 0.99-7.93)]. More experienced panel members tended to issue weaker recommendations [OR=0.891 (95CI% 0.795 to 0.98)] (see Discussion), while the tendency to employ a maximizing cognitive style when faced with decision difficulties was associated with OR=2.14(95% CI .99-4.65) (Table 3).

Table 4 outlines the analysis when the panel members issued recommendations ***“strong against” vs. “weak against”*** health interventions (Fig 3). In this analysis, only one formal GRADE factor (the importance of patients' V&P) had an effect, while 5 non-GRADE factors displayed statistically significant association. More experienced panel members, those with higher intolerance of uncertainty and those with propensity toward analytical thinking tended to issue weaker recommendations against the intervention. On other hand, being a methodologist, older, recused from voting due to a conflict of

interest, or issuing guidelines for a vulnerable population were associated with strong recommendation against intervention.

Agreement between individual panel members and the group regarding strength of recommendations (SOR) ranged from poor (kappa ranging from: -0.01 to 0.03; 2 panels) to fair (kappa range: 0.21 to 0.47; 4 panels) to moderate (kappa=0.64; 1 panel) (Fig 4). Agreement of judgments related to voting “for” and “against” the intervention was somewhat better 0.37 (95%CI 0.16 to 0.58) and 0.42 (95%CI: 0.19 to 0.64) respectively (data not shown).

Finally, we calculated ICC to determine the extent of the overall variation in the response of the panel members. The results varied with the analyses: in the main analysis (Tables 2) we found negligible correlation between individual vote and the panel voting pattern (ICC=0.06) but in the analysis that omitted the recommendations in which the panel didn’t issue a recommendation determining the strength of association and direction of the vote, ICC was 0.50 in recommendations “for” and 0.48 in recommendations “against”.

## **Discussion**

We report the first study evaluating the impact of the GRADE system, and non-GRADE factors that could impact on guidelines panel members’ decision making. Overall, we showed that factors associated with GRADE’s conceptual framework were, in general, highly associated with SOR. A secondary analysis suggested, however, that certainty of evidence may have little or no influence on SOR when a panel makes recommendations against an intervention. We also detected statistical association

between SOR and non-GRADE factors but, aside from age/clinical experience, these varied across statistical models.

The main findings likely reflect the effect of instructions<sup>29,30</sup> due to use of the highly structured GRADE EtD framework.<sup>31</sup> Adherence to structure is typically seen with high-ability participants (such as expert panelists) who can follow instructions that require cognitive effort and suppress the influence of other factors and prior beliefs.<sup>29,32</sup> The findings extend the observations from our qualitative analysis<sup>33</sup> that policy-makers and users of guidelines who apply GRADE methods may expect that the guideline panels will not only rely on GRADE factors but use the cognitive processes that facilitate decision making according to the (GRADE) instructions. Nevertheless, individual characteristics such as age, experience, intolerance of uncertainty, and propensity toward analytical thinking were also, in some models, associated with SOR. Theoretically, a type of a task and instructions can activate cognitive processes to align them toward accomplishing stated goals.<sup>34</sup> For example, the importance of intolerance of uncertainty can be seen as a response to the underlying clinical uncertainties that activate analytical reasoning processes that prompted development of guidelines in the first place.<sup>35</sup>

Our results regarding the importance of certainty of evidence is consistent with observations in two smaller studies.<sup>36,37</sup> The results provide empirical verification of the key EBM normative principle regarding the relationship between the credibility of underlying evidence and willingness to endorse a health intervention<sup>1</sup>: when certainty of evidence is high, we can expect that most panelists will issue strong recommendations.

However, this relationship disappeared when the panel members “voted” “against” health intervention. Potential explanations for this finding include: 1) the “yes/for” bias<sup>25,26,38-40</sup>: according to dual process theory of acquiescence people are overall slower to respond to “no” than to “yes”. “Yes” (“for”) responses tap into “*feeling of rightness*” *heuristic* that the answer is correct: is automatic, *effortless* (type 1 process), which is activated much faster than *effortful* (type 2 processes) associated with processing of “no” (“against”) responses.<sup>38-40</sup>; 2) Voting “against” an intervention is cognitively more challenging because people need to mentally simulate the consequences of two contradictory assessments- certainty of evidence, which moves from very low to high in “positive” direction and strength of recommendation “against” the intervention, which goes in the opposite direction. This often occurs when cognitive resources are depleted<sup>41,42</sup> as when people are tired and decision-making occurs in time-constraint settings, which characterize most human engagements including guidelines development process; 3) in a number of cases, the question was formulated as a “vote” against intervention without explicit description of a comparator, which may have introduced a reference class problem (i.e., when reference category is not well specified, people’s estimates are often incorrect)<sup>43,44</sup>; 4) GRADE paradigmatic situations that justify strong recommendations despite low certainty evidence, may have occurred more in recommendations “against” than “for” interventions.<sup>45</sup>

As in all research, we cannot exclude the possibility that some associations we observed may be simply due to chance. For example, as in our earlier study,<sup>20</sup> we detected that effect of age and experience went in the opposite direction, which we judged to be a spurious association. This occurred because in medicine, as in many

professions, age and experience are positively correlated ( $r=0.85$  in this study) across individuals, making it difficult to isolate the unique influence of a given variable on the third variable.

Our results also provide empirical support for the importance of managing conflict of interests<sup>46,47</sup> - the panel members who were required to recuse themselves, had they been allowed to vote, would have registered different views from those of their colleagues.

The frequent low agreement between judgments of individual panel members' and the group consensus related to SOR raises the possibility that the apparent consensus represents individual panel members' conforming to the group<sup>48,49</sup>, particularly since more than 50% of discussion was dominated by chairs and co-chairs.<sup>33</sup> In a classic paper on opinions and social pressure, Asch warned that *"Consensus is an indispensable condition in a complex society, but consensus, to be productive, requires that each individual contribute independently out of experience and insight. When consensus is produced by conformity, the social process is polluted"*.<sup>48</sup>

Nevertheless, fewer than 2% of participants (Table 2) reported that they felt any pressure to conform to the group vote. Earlier studies suggested that when instructions that clearly operationalize procedures are provided, agreement on assessments such as the certainty of evidence becomes high.<sup>50</sup> Lack of familiarity with the GRADE system (despite introductory lectures about GRADE) and the complexity of the judgement inherent in making recommendations may explain the low agreement we observed. An alternative explanation is that many of the decisions were close calls in the panels where agreement was low – and fewer when agreement was high.

Another explanation for low agreement arises from our observation that variability in V&P was associated with SOR despite, as we have reported previously, only 1% of the discussion was devoted to this issue<sup>33</sup>. It is possible that panelists had different views of the extent of diversity and uncertainty in V&P, views that they did not express in group discussion. This suggests that chairs of guideline using GRADE should insist on repeated discussion of V&Ps issues.

### *Strengths and Limitations*

A strength of our study is that it is the first to assess the decision making of guidelines panels in natural, real-life setting. At the same time, the observational design precluded experimental control of the variables that may allow drawing stronger inferences. Hence, future studies will be necessary to establish the generalizability of our findings. Nevertheless, our findings represent first initial insights into how guideline decision-making works in real life, and suggests possible improvements in the process

In conclusion, we found that policy-makers and users of guidelines who apply GRADE methods may expect that the guideline panels will rely on GRADE factors. However, low agreement between individual panel members and group consensus suggests that the process can be improved, perhaps by further operationalization of GRADE criteria, by better training of the panel members in GRADE methodology, and by framing, as far as is possible, all recommendations in terms of voting “for” instead of “against” a given health intervention.

**Acknowledgment:** This project was supported by grant number R01HS024917 from the Agency for Healthcare Research and Quality. The content is solely the responsibility of the authors and does not necessarily represent the official views of the Agency for Healthcare Research and Quality. PI: Dr. Djulbegovic. We thank Robert Kunkle and his staff from the American Society of Hematology (ASH) for helping facilitate our project and Robert Hamm of the University of Oklahoma Health Sciences Center for useful feedback on an earlier version of the manuscript. We also wish to thank all panelists participating in development of the ASH guidelines on venous thromboembolism and immune thrombocytopenia for agreeing to take part in our study. We also thank the ASH leadership for supporting this project.

**Role of Funding:** Funding sources have no involvement in any of the activities conducted in this article.



## Legend:

Fig 1 Overview of data collection process. Data were collected from the guidelines panels convened by The American Society of Hematology (ASH) to develop guidelines for the management of the following conditions: (a) prevention of venous thromboembolism (VTE) in surgical hospitalized patients, (b) prevention of VTE in medical patients, (c) heparin-induced thrombocytopenia (HIT), (d) management of thrombophilia, (e) VTE in the context of pregnancy, (f) VTE in pediatric populations, (g) optimal management of anticoagulation therapy, (h) VTE in patients with cancer, (i) treatment of VTE, and (j) management of immune thrombocytopenia (ITP).

Unfortunately, data collection from two panels were not recorded due to technical glitches. The final analysis included data from 8 panels (see text)

Fig 2. Effect of certainty of evidence and judgments about the balance of benefits and harms (in favor of intervention over comparator). The vertical line around the each point denotes a 95% confidence interval.

Fig 3. A relationship between the quality (certainty) of underlying evidence and the probability of issuing of a strong recommendation FOR (a) vs. AGAINST (b) a given health intervention. The vertical line around each point denotes a 95% confidence interval. The results remained the same even though panelists were instructed to align strength of recommendations with direction of recommendations (the reminders were originally issued orally, but it was included in the survey for the last 5 panels)

Fig 4. Agreement in judgements related to strength of recommendations between individual panel members and the group judgements

## References

1. Djulbegovic B, Guyatt GH. Progress in evidence-based medicine: a quarter century on. *Lancet*. 2017;390(10092):415-423.
2. Graham R, Mancher M, Wolman DM, Greenfield S, Steinberg E, Editors. *Clinical practice guidelines we can trust*. Washington, DC: Institute of Medicine, National Academies Press; 2011.
3. Vandvik P, Brandt L, Alonso-Coello P, Treweek S, Akl E, Kristiansen A. Creating clinical practice guidelines we can trust, use, and share: a new era is imminent. *Chest*. 2013;144:381 - 389.
4. Keeney R. Personal decisions are the leading cause of death. *Operations Research* 2008;56(6):1335–1347.
5. Medicine) Ilo. *Variation in health care spending: Target decision making, not geography*. . Washington, DC:: The National Academies Press.;2013.
6. Pronovost PJ. Enhancing physicians' use of clinical guidelines. *JAMA*. 2013;310(23):2501-2502.
7. Djulbegovic B. A Framework to Bridge the Gaps Between Evidence-Based Medicine, Health Outcomes, and Improvement and Implementation Science. *Journal of Oncology Practice*. 2014;10(3):200-202.
8. Rosenthal MB. Physician Payment after the SGR — The New Meritocracy. *New England Journal of Medicine*. 2015;373(13):1187-1189.
9. Berkman ND LK, Ansari M, McDonagh M, Balk E, Whitlock E, Reston J, Bass E, Butler M, Gartlehner G, Hartling L, Kane R, McPheeters M, Morgan L, Morton SC, Viswanathan M, Sista P, Chang S. AHRQ Publication No. 13(14)-EHC130-EF. . *Grading the Strength of a Body of Evidence When Assessing Health Care Interventions for the Effective Health Care Program of the Agency for Healthcare Research and Quality: An Update. Methods Guide for Comparative Effectiveness Reviews (Prepared by the RTI-UNC Evidence-based Practice Center under Contract No. 290-2007-10056-I)*. Rockville, MD: Agency for Healthcare Research and Quality. November 2013.;2013.
10. Ebell MH, Siwek J, Weiss BD, et al. Strength of recommendation taxonomy (SORT): a patient-centered approach to grading evidence in the medical literature. *Am Fam Physician*. 2004;69(3):548-556.
11. Harbour R, Miller J, Group fSIGNGR. A new system for grading recommendations in evidence-based guidelines. *BMJ*. 2001;323:334-336.
12. Petitti DB, Teutsch SM, Barton MB, et al. Update on the Methods of the U.S. Preventive Services Task Force: Insufficient Evidence. Vol 1502009:199-205.
13. Hill J, Bullock I, Alderson P. A Summary of the Methods That the National Clinical Guideline Centre Uses to Produce Clinical Guidelines for the National Institute for Health and Clinical Excellence. *Annals of Internal Medicine*. 2011;154(11):752-757.
14. Guyatt GH, Oxman AD, Vist GE, et al. GRADE: an emerging consensus on rating quality of evidence and strength of recommendations. *Bmj*. 2008;336(7650):924-926.

15. GRADE Working Group. Organizations that have endorsed or that are using GRADE. 2016; <http://www.gradeworkinggroup.org/society/index.htm>. Accessed April 21, 2016.
16. Alonso-Coello P, Oxman AD, Moberg J, et al. GRADE Evidence to Decision (EtD) frameworks: a systematic and transparent approach to making well informed healthcare choices. 2: Clinical practice guidelines. *BMJ*. 2016;353.
17. Alonso-Coello P, Schünemann HJ, Moberg J, et al. GRADE Evidence to Decision (EtD) frameworks: a systematic and transparent approach to making well informed healthcare choices. 1: Introduction. *BMJ*. 2016;353.
18. Appelt KC, Milch KF, Handgraaf MJJ, Weber EU. The Decision Making Individual Differences Inventory and guidelines for the study of individual differences in judgment and decision-making research. *Judgment and Decision Making*. 2011;6(252–262).
19. Hastie R, Dawes RM. *Rational choice in an uncertain world. 2nd edition*. Los Angeles: Sage Publications, Inc.; 2010.
20. Djulbegovic B, Beckstead JW, Elqayam S, et al. Evaluation of Physicians' Cognitive Styles. *Med Decis Making*. 2014;34(5):627-637.
21. Djulbegovic M, Beckstead J, Elqayam S, et al. Thinking Styles and Regret in Physicians. *PLoS One*. 2015;10(8):e0134038.
22. Budner S. Intolerance of ambiguity as a personality variable. *J Pers*. 1962;30:29-50.
23. Turner BM, Rim HB, Betz NE, Nygren TE. The Maximization Inventory *Judgment and Decision Making*. 2012;7(1):48-60.
24. Pacini R, Epstein S. The relation of rational and experiential information processing styles to personality, basic beliefs, and the ratio-bias phenomenon *Journal of Personality and Social Psychology*. 1999;76:972-987.
25. Gilbert DT. How Mental Systems Believe. *American Psychologist*. 1991;46(2):107-119.
26. Tversky A, Koehler DJ. Support theory: a nonextensional representation of subjective probability. *Psychol Rev*. 1994;101:547-567.
27. Sun S. Meta-analysis of Cohen's kappa. *Health Serv Outcomes Res Method*. 2011;11:145–163.
28. STATA, ver. 14 [computer program]. College Station, TX 2013.
29. Evans JS. Logic and human reasoning: an assessment of the deduction paradigm. *Psychol Bull*. 2002;128(6):978-996.
30. Heit E, Rotello CM. Traditional difference-score analyses of reasoning are flawed. *Cognition*. 2014;131(1):75-91.
31. Guyatt GH, Oxman AD, Kunz R, et al. Going from evidence to recommendations. *Bmj*. 2008;336(7652):1049-1051.
32. Evans JSBT, Handley SJ, Neilens H. The influence of cognitive ability and instructional set on causal conditional inference. *THE QUARTERLY JOURNAL OF EXPERIMENTAL PSYCHOLOGY*. 2010;63(5):892-909.
33. Li SA, Alexander PE, Reljic T, et al. Evidence to Decision framework provides a structured “roadmap” for making GRADE guidelines recommendations. *J Clin Epidemiol*. 2018;104:103-112.

34. Phillips WJ, Fletcher JM, Marks AD, Hine DW. Thinking styles and decision making: A meta-analysis. *Psychol Bull.* 2016;142(3):260-290.
35. Djulbegovic B, Hozo I, Greenland S. Uncertainty in Clinical Medicine. In: Gifford F, ed. *Philosophy of Medicine (Handbook of the Philosophy of Science)*. London: Elsevier; 2011:299-356.
36. Djulbegovic B, Trikalinos TA, Roback J, Chen R, Guyatt G. Impact of quality of evidence on the strength of recommendations: an empirical study. *BMC Health Serv Res.* 2009;9(1):120.
37. Djulbegovic B, Kumar A, Kaufman RM, Tobian A, Guyatt GH. Quality of evidence is a key determinant for making a strong guidelines recommendation. *J Clin Epidemiol.* 2015;68(7):727-732.
38. Thompson VA, Evans JSBT, Campbell JID. Matching bias on the selection task: It's fast and feels good. *Thinking & Reasoning.* 2013;19(3):431-452,.
39. Shynkaruk JM, Thompson VA. Confidence and accuracy in deductive reasoning. *Memory & Cognition.* 2006;34(3):619-632.
40. Knowles ES, Condon CA. Why people say "yes": A dual-process theory of acquiescence. . *Journal of Personality and Social Psychology.* 1999;77(2):379-386.
41. De Neys W. Dual processing in reasoning: two systems but one reasoner. *Psychol Sci.* 2006;17(5):428-433.
42. De Neys W, Verschueren N. Working memory capacity and a notorious brain teaser: the case of the Monty Hall Dilemma. *Exp Psychol.* 2006;53(2):123-131.
43. Gigerenzer G. Content-blind norms, no norms, or good norms? A reply to Vranas. *Cognition.* 2001;81:93-103.
44. Gigerenzer G. What are natural frequencies? *BMJ.* 2011;343.
45. Andrews JC, Schunemann HJ, Oxman AD, et al. GRADE guidelines 15: Going from evidence to recommendation-determinants of a recommendation's direction and strength. *J Clin Epidemiol.* 2013;66(7):726-735.
46. George JN, Vesely SK, Woolf SH. Conflicts of interest and clinical recommendations: comparison of two concurrent clinical practice guidelines for primary immune thrombocytopenia developed by different methods. *Am J Med Qual.* 2014;29(1):53-60.
47. Institute of Medicine (US) Committee on Conflict of Interest in Medical Research E, and Practice; Lo B, Field MJ, editors,. *Conflict of Interest in Medical Research, Education, and Practice*. Washington (DC): : National Academies Press (US); 2009.
48. Asch SE. Opinions and Social Pressure. *Sci American.* 1955;193(5):31-35.
49. Asch SE. Studies of independence and conformity. I A minority of one against unanimous majority. *Psychological Monographs: General and Applied.* 1956;70(9):1-70.
50. Kumar A, Miladinovic B, Guyatt GH, Schünemann H, Djulbegovic B. GRADE guidelines system is reproducible when instructions are clearly operationalized even among the guidelines panel members with limited experience with GRADE *J Clin Epidemiol.* 2016;75:115-118.

## What is new?

### Key findings:

- The GRADE guidelines system specifies factors that guidelines panels **should** take into considerations when issuing recommendations. However, many other (non-GRADE) factors may also affect recommendations.
- To what extent GRADE vs. non-GRADE factors influence guidelines panels' decision-making remains uncertain.
- We found that GRADE factors affect guidelines decision-making process more than non-GRADE factors, likely due to the effect of instructions provided within structured GRADE Evidence-to-Decision (EtD) framework. Consistent with principles of evidence-based medicine, we confirmed relation between the certainty of evidence and strength of recommendations.
- The findings remained robust when panels issued recommendations **for** health interventions. However, when the panels generated recommendations **against** health interventions, the relation between certainty of evidence and strength of recommendations disappeared pointing to the existence of so called "yes" bias (people acquiesce to "yes" statements more readily than to "no" statements)
- Even within highly structured GRADE process, the panel members demonstrated **variability** in their **individual** responses (kappa between individual panel members and the group consensus vote for strength of recommendations ranged from very poor (-0.01) to moderate (0.64)),
- Depending on the analytical model, some non-GRADE factors were also associated with the strength of recommendations issued by the panels. Different non-GRADE factors were associated with recommendations "for" vs. "against" health interventions. However, age/clinical experience of the panelists remained statistically significant across all models.

### What this adds to what is known:

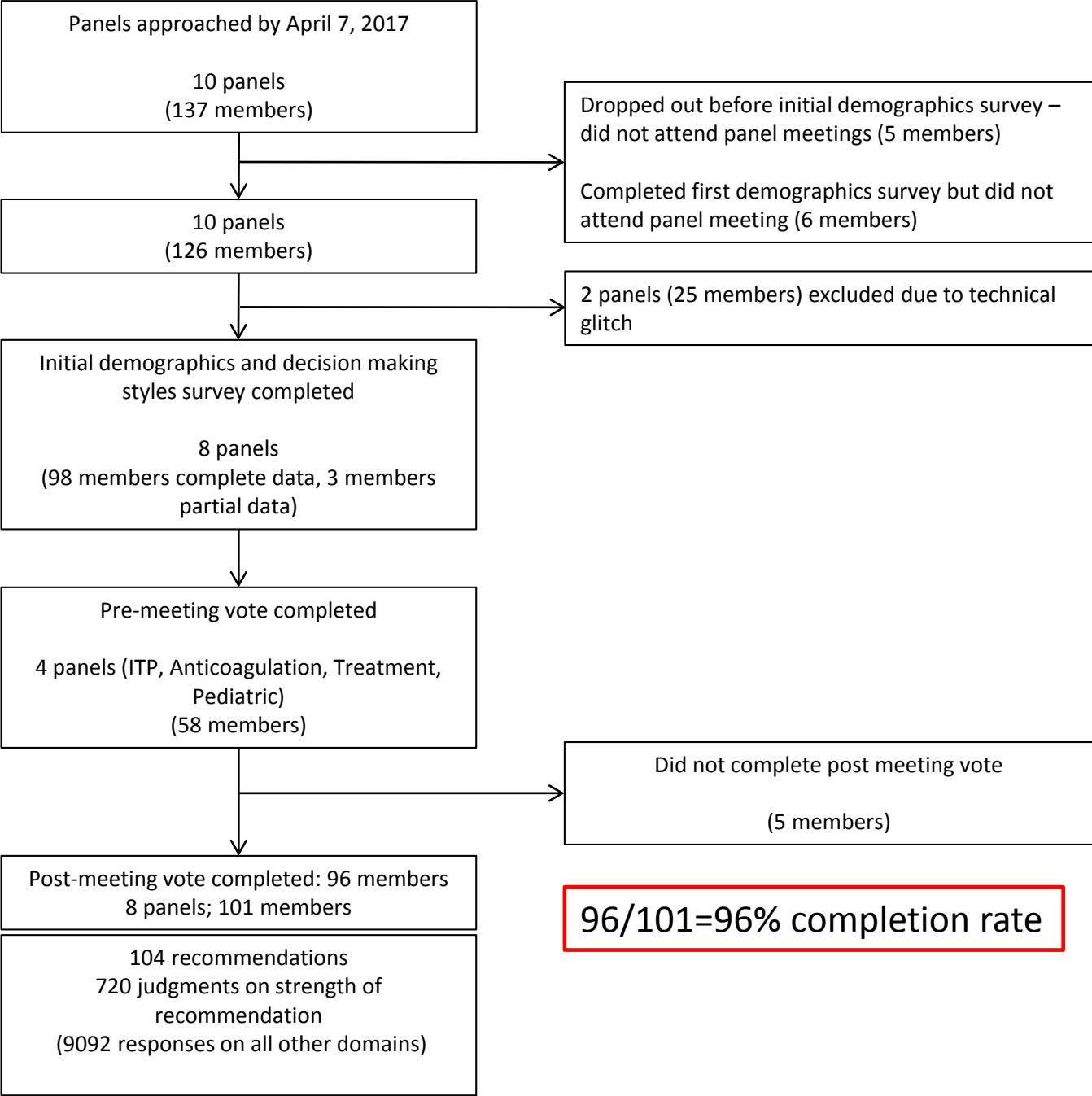
- This quantitative analysis of 8 panels confirms that GRADE instruction given within EtD structured framework results in consideration of GRADE factors as intended by the GRADE system.
- The system does not, however, appear to give consistent results when the panels issue recommendation **for** vs **against** health intervention.
- In addition, individual member "assessment" often considerably differ from the group, consensus vote.

### What is the implication, and what should change now:

- Guideline panels that place a high value on adherence to the GRADE system should consider use of EtD framework in developing their recommendations.
- To avoid "yes" bias, guidelines developers should, in most instances, express all recommendations as a vote "for" instead of "against" recommendations

- Exploration of reasons why panel members are sometimes in agreement and sometimes not may inform the need for additional strategies such as more extensive training in GRADE to reduce variability.

Figure 1. Participant recruitment for evaluation of clinical guidelines development process ASH panels



# Effect of certainty of evidence and judgments about the balance of benefits and harms (in favor of intervention over comparator)

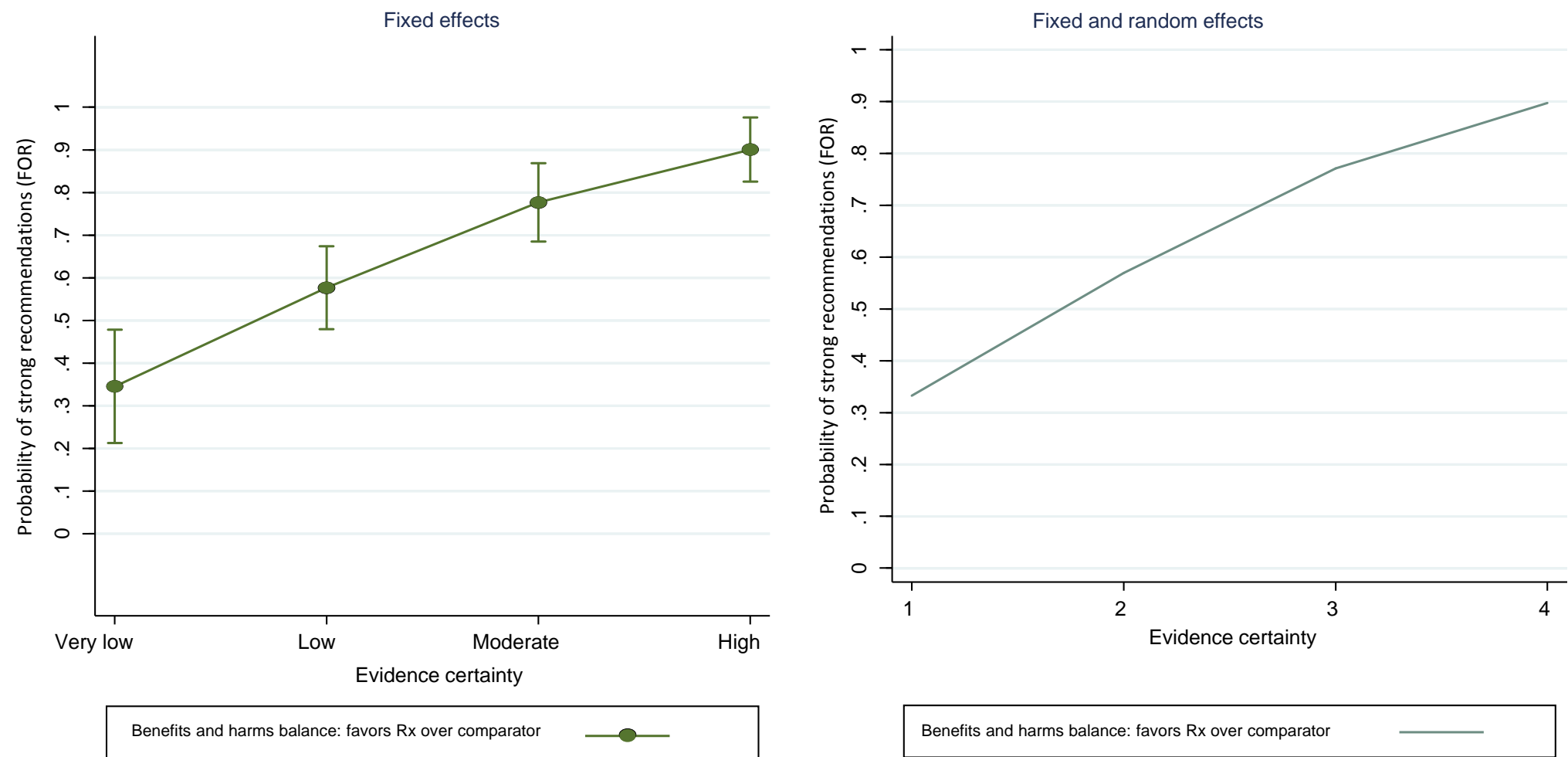


Fig 2



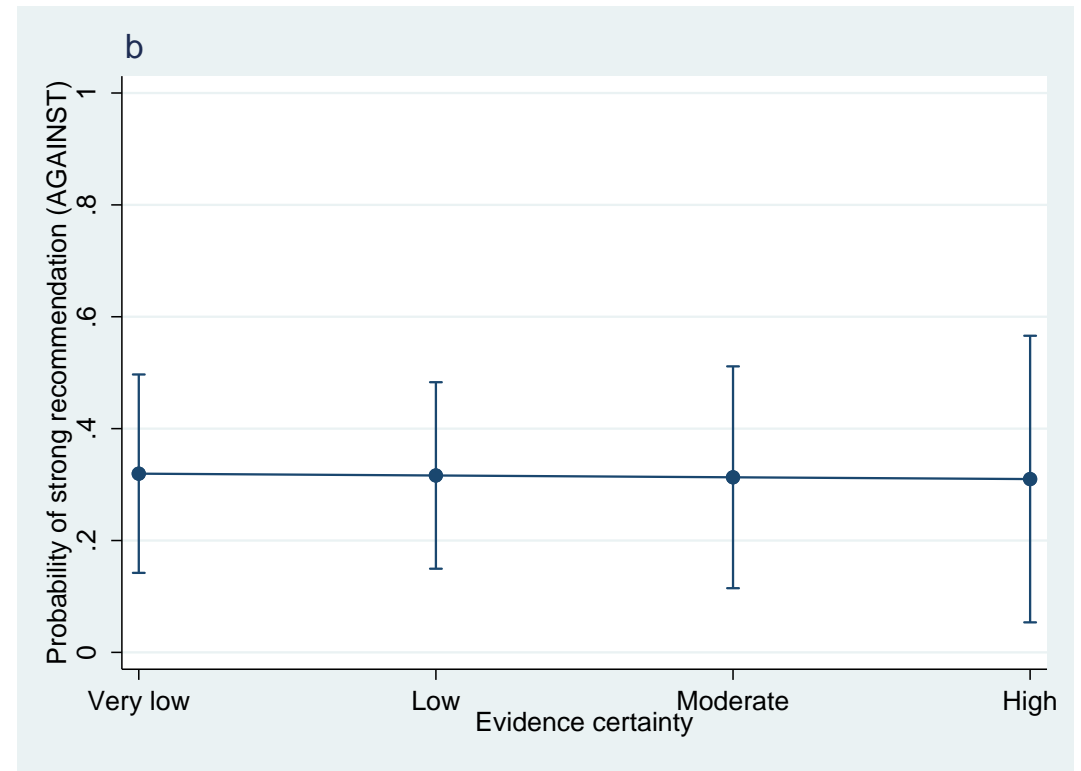
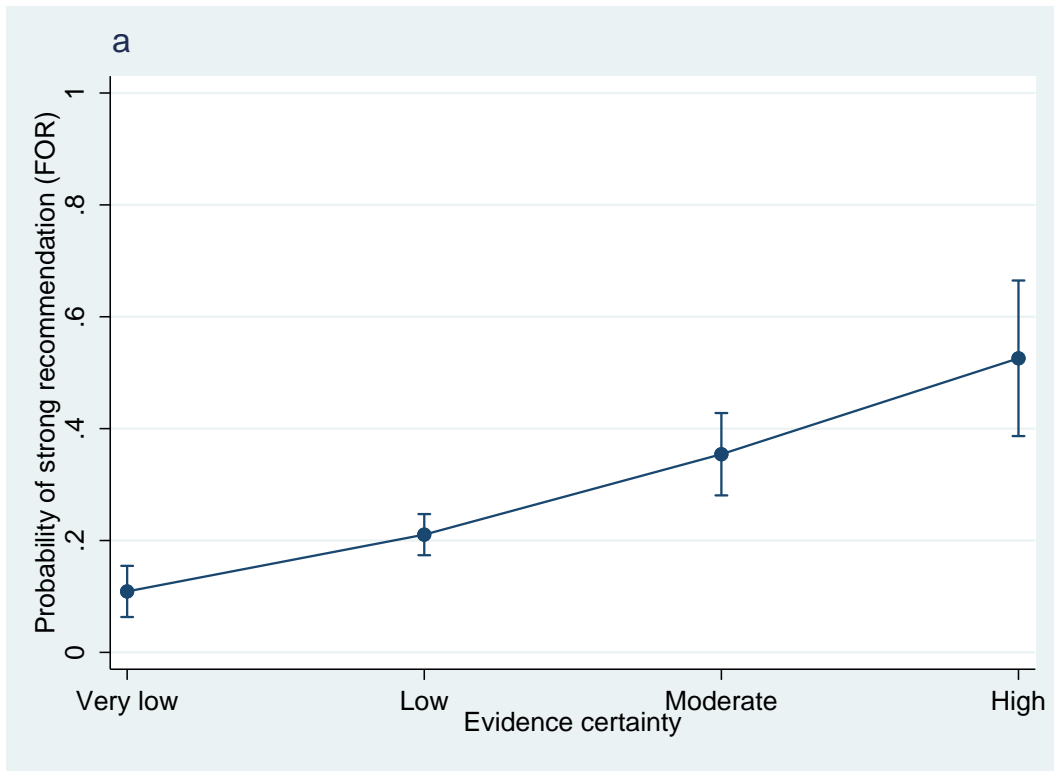
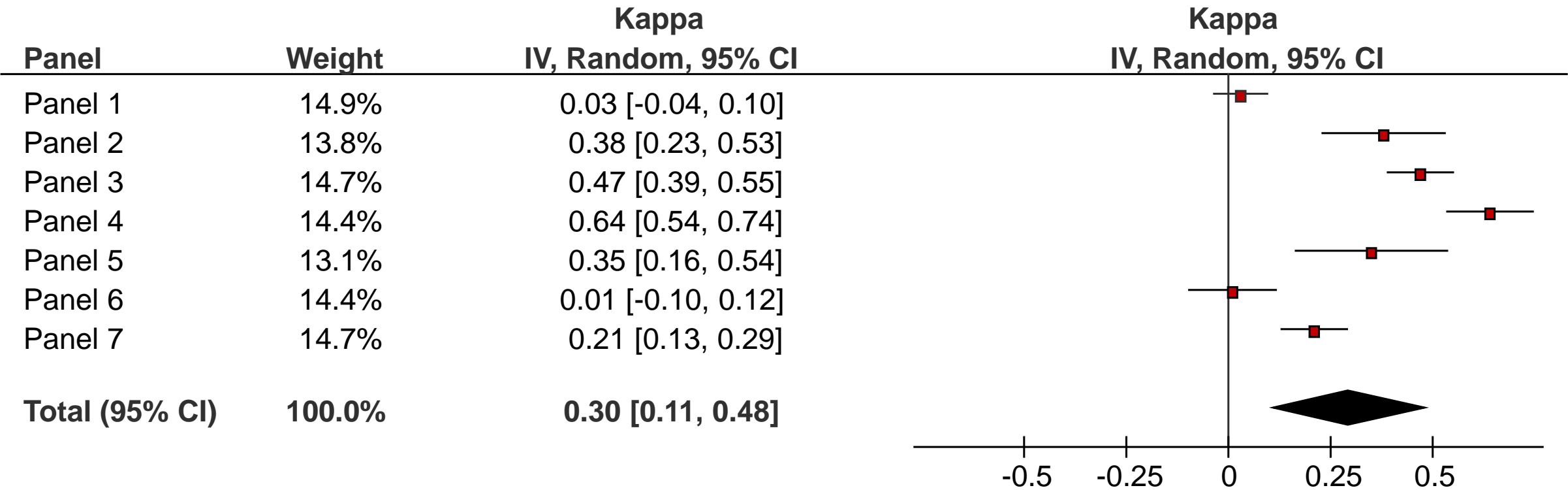


Fig 3

Agreement in judgements related to strength of recommendations between individual panel members and the group judgements



Heterogeneity:  $\text{Tau}^2 = 0.06$ ;  $\text{Chi}^2 = 150.70$ ,  $\text{df} = 6$  ( $P < 0.00001$ );  $I^2 = 96\%$   
Test for overall effect:  $Z = 3.16$  ( $P = 0.002$ )

Fig 4

**Table 1.** Participant characteristics by panel<sup>#</sup>

Variable	Overall N (%)
Number of participants	101
Age, median (quartile1, quartile 3); range	48.0 (41, 56) (28 to 78)
Sex	
Male	56 (55.4)
Female	45 (44.6)
Role	
Chair	8 (7.9)
Methodologist	16 (15.8)
Clinician	66 (65.4)
Patient representatives	11 (10.9)
Panel	
Anticoagulation	13 (12.9)
Cancer	15 (14.9)
Heparin-Induced Thrombocytopenia	11 (10.9)
Immune thrombocytopenia	16 (15.8)
Pediatric	15 (14.9)
Pregnancy	10 (9.9)
Thrombophilia	7 (6.9)
Treatment	14 (13.9)
Country of origin	
United States	52 (51.5)
Canada	25 (24.8)
Netherlands	5 (5.0)
Italy	3 (3.0)
United Kingdom	3 (3.0)
Germany	3 (3.0)
Australia	3 (3.0)
Austria	2 (2.0)
Argentina	1 (1.0)
Belgium	1 (1.0)
Denmark	1 (1.0)
New Zealand	1 (1.0)
Switzerland	1 (1.0)
Years of experience*, median (quartile1, quartile 3); (range)	18 (11,26) (2 to 49)
Self-reported level of experience*	
Higher than others	46 (55.4)
About same as others	32 (38.6)
Lower than others	5 (6.0)
How many patients with similar condition do you treat per month*	
None	13 (16.3)
1 to 5	14 (17.5)
6 to 10	3 (3.8)
11 to 15	7 (8.8)
More than 15	43 (53.8)

<sup>#</sup>included in the final analysis; there was no statistically significant difference between these participants and those that were excluded from the analysis (see Fig 1)\* Questions regarding professional experience were only answered by clinicians

Table 2 Association Between Decision Making Factors and the Strength of Recommendations	
Mixed-effect model [ordered logistic regression]	Dependent Variable: Neither For/Against; Weak For/Weak Against; Strong For/ Strong Against
FIXED EFFECT	Odds Ratio (OR) with 95% Confidence Interval (CI)
Role: chair (reference category) methodologist patient representative panel member	1.47(95% CI .57-3.80;p=.43 0.38(95% CI .13-1.18;p=.09) 1.07(95% CI .58-1.98;p=.83)
Vulnerable population Yes vs. no	1.27(95% CI .81-2.00;p=.3)
Pressured to "vote" certain way	0.84(95% CI .31-2.31;p=.74)
Recused from "voting" Yes vs. no	1.36(95% CI .92-2.03;p=.13)
<b>Age (per decade)</b>	<b>1.79(95% CI 1.2 to 2.84;p=0.005)</b>
Sex Female vs. male	1.10(95% CI .75-1.62;p=.61)
Experience (years in management of given condition)	0.97(95% CI .926-1.00;p=.09)
Expertise (considers oneself with higher, same or low expertise than most other experts)	0.74(95% CI .51-1.09;p=.13)
Exposure (# of patients per month with given condition)	0.90(95% CI .770-1.06;p=.21)
Objectivism (tendency to seek empirical information)	1.29(95% CI .78-2.17;p=.34)
Tendency toward rational (analytical) thinking	0.66(95% CI .36-1.10;p=.11)
Tendency toward experiential-intuitive thinking	0.95(95% CI .67-1.34;p=.76)
Satisficing (tendency to accept "good" enough solution)	1.13(95% CI .63-2.02;p=.68)
Maximizing(decision difficulty)-degree difficulty experienced when making choices among abundant options	1.05(95% CI .78-1.41;p=.766)
Maximizing(alternative search)-tendency to expand resources in search for best possible solution	1.08(95% CI .81-1.43;p=.606)
<b>Intolerance of uncertainty</b>	<b>0.57(95% CI .37-.86;p=0.008)</b>
Regret of making a wrong recommendation	0.99(95% CI .98-1.02;p=.93)
<b>Certainty in Evidence</b>	<b>1.84(95% CI 1.46-2.31;p&lt;0.0001)</b>
<b>Importance of patients' values and preferences</b>	<b>1.48(95% CI 1.15-1.89;p=0.002)</b>
<b>Balance between benefits and harms</b>	<b>1.49(95% CI 1.31-1.70;p&lt;0.0001)</b>
Importance of cost and resources	1.06(95% CI .86-1.28;p=.56)
<b>RANDOM INTERCEPTS</b>	
Panel (variance)	0.62(95% CI .19-2.05)
Participant within panel (variance)	2.18*10 <sup>-34</sup>

Table 3 Association Between Decision Making Factors and the Strength of Recommendations	
Mixed-effect model [logistic regression]	Dependent Variable: <b>Weak For; Strong For</b> Odds Ratio (OR) with 95% Confidence Interval (CI)
<b>FIXED EFFECT</b>	
Role: chair (reference category) <b>methodologist</b> patient representative panel member	<b>0.06 (95% CI 0-.85;p=.04)</b> 0.64(95% CI .03-12.4;p=.77) 1.15(95% CI .23-5.89;p=.87)
Vulnerable population Yes vs. no	1.83(95% CI .56-5.91;p=.31)
Pressured to "vote" certain way	2.27(95% CI .21-24.9;p=.5)
Recused from "voting" Yes vs. no	1.18(95% CI .4-3.45;p=.76)
<b>Age (per decade)</b>	<b>2.6 (95% CI .99-7.93;p=.079)</b>
Sex Female vs. male	0.55(95% CI .20-1.48;p=.24)
<b>Experience (years in management of given condition)</b>	<b>0.89(95% CI .76-.99;p=.047)</b>
Expertise (consider oneself with higher, same or low expertise than most other experts)	1.40(95% CI .566-3.49;p=.47)
Exposure (# of patients per month with given condition)	0.89(95% CI .61-1.31;p=.57)
Objectivism (tendency to seek empirical information)	1.64(95% CI .44-6.10;p=.46)
Tendency toward rational (analytical) thinking	0.51(95% CI .15-1.76;p=.23)
Tendency toward experiential-intuitive thinking	0.89(95% CI .37-2.15;p=.79)
Satisficing (tendency to accept "good" enough solution)	2.23(95% CI .61-8.1;p=.23)
<b>Maximizing(decision difficulty)-degree difficulty experienced when making choices among abundant options</b>	<b>2.14(95% CI .99-4.65;p=.05)</b>
Maximizing(alternative search)-tendency to expand resources in search for best possible solution	0.65(95% CI .31-1.35;p=.25)
Intolerance of uncertainty	0.4(95% CI .13-1.21;p=.11)
Regret of making a wrong recommendation	0.99(95% CI .95-1.05;p=.89)
<b>Certainty in Evidence</b>	<b>3.61(95% CI 2.17-6.01;p&lt;0.001)</b>
<b>Importance of patients' values and preferences</b>	<b>2.33(95% CI 1.34-4.07;p=.003)</b>
<b>Balance between benefits and harms</b>	<b>18.3(95% CI 7.68-43.7;p=.000)</b>
<b>Importance of cost and resources</b>	<b>1.83(95% CI 1.25-2.67;p=.002)</b>
<b>RANDOM INTERCEPTS</b>	
Panel (variance)	3.14*10 <sup>-32</sup>
Participant within panel (variance)	1.00*10 <sup>-33</sup>

<b>Table 4 Association Between Decision Making Factors and the Strength of Recommendations</b>	
Mixed-effect model [logistic regression]	<b>Dependent Variable: <u>Weak Against; Strong Against</u> Odds Ratio (OR) with 95% Confidence Interval (CI)</b>
<b>FIXED EFFECT</b>	
Role: chair (reference category) methodologist patient representative panel member	<b>11.7(95% CI 1.50-91.4;p=.019)</b> 1.28(95% CI .070-23.4;p=.87) 2.34(95% CI .558-9.83;p=.25)
<b>Vulnerable population Yes vs. no</b>	<b>3.97(95% CI 1.30-12.1;p=.015)</b>
Pressured to "vote" certain way	0.34(95% CI .036-3.28;p=.35)
<b>Recused from "voting" Yes vs. no</b>	<b>3.34(95% CI 1.29-8.65;p=.013)*</b>
<b>Age (per decade)</b>	<b>4.8(95% CI 1.8-12.76;p=.002)</b>
Sex Female vs. male	0.56(95% CI .247-1.43;p=.25)
<b>Experience (years in management of given condition)</b>	<b>0.87(95% CI .79-.96;p=.007)</b>
Expertise (consider oneself with higher, same or low expertise than most other experts)	0.46(95% CI .18-1.15;p=.09)
Exposure (# of patients per month with given condition)	0.84(95% CI .58-1.20;p=.33)
Objectivism (tendency to seek empirical information)	1.15(95% CI .28-4.63;p=.84)
<b>Tendency toward rational (analytical) thinking</b>	<b>0.215(95% CI .063-.729;p=.014)</b>
Tendency toward experiential-intuitive thinking	1.32(95% CI .60-2.90;p=.49)
Satisficing (tendency to accept "good" enough solution)	1.31(95% CI .36-4.72;p=.68)
Maximizing(decision difficulty)-degree difficulty experienced when making choices among abundant options	1.45(95% CI .74-2.85;p=.28)
Maximizing(alternative search)-tendency to expand resources in search for best possible solution	1.00(95% CI .55-1.84;p=.99)
<b>Intolerance of uncertainty</b>	<b>0.16(95% CI .05-.49;p=.001)</b>
Regret of making a wrong recommendation	OR=1.01(95% CI .97-1.06;p=.62)
Certainty in Evidence	OR=.98(95% CI .57-1.68;p=.94)
<b>Importance of patients' values and preferences</b>	<b>2.26(95% CI 1.32-3.87;p=.003)</b>
Balance between benefits and harms	0.78(95% CI .52-1.23;p=.31)
Importance of cost and resources	1.01(95% CI .60-1.69;p=.97)
<b>RANDOM INTERCEPTS</b>	
Panel (variance)	1.54(95% CI .380-6.23)
Participant within panel (variance)	1.47*10 <sup>-36</sup>

\*when interactions with the certainty of evidence was taken into account, OR= 4.63 (95% CI 0.814 to 26.38; p=0.084)

**Conflict of Interest Statement**

Although most authors have worked on development of GRADE system, we declare no conflict of interest related to the content and analysis of this paper.

## **Author Statement**

BD has conceptualized the study, received funding and wrote the first draft. GG helped with the grant proposal and revised the first draft of the paper, which has then be shared and approved by all authors. TR, SAL,PA,RN, WW have helped with data collection. HS and AC have helped with the study logistics and provided the intellectual input from the guidelines process perspective. SE provided the perspective from psychology of decision-making. IH and QZ helped with data analysis. BD serves as a guarantor.



## Appendix 1. GRADE Evidence to Decision Framework template

### Assessment

	JUDGEMENT	RESEARCH EVIDENCE	ADDITIONAL CONSIDERATIONS
PROBLEM	<b>Is the problem a priority?</b> <ul style="list-style-type: none"> <li>○ No</li> <li>○ Probably no</li> <li>○ Probably yes</li> <li>○ Yes</li> <li>○ Varies</li> <li>○ Don't know</li> </ul>		
DESIRABLE EFFECTS	<b>How substantial are the desirable anticipated effects?</b> <ul style="list-style-type: none"> <li>○ Trivial</li> <li>○ Small</li> <li>○ Moderate</li> <li>○ Large</li> <li>○ Varies</li> <li>○ Don't know</li> </ul>		
UNDESIRABLE EFFECTS	<b>How substantial are the undesirable anticipated effects?</b> <ul style="list-style-type: none"> <li>○ Large</li> <li>○ Moderate</li> <li>○ Small</li> <li>○ Trivial</li> <li>○ Varies</li> <li>○ Don't know</li> </ul>		
CERTAINTY OF EVIDENCE	<b>What is the overall certainty of the evidence of effects?</b> <ul style="list-style-type: none"> <li>○ Very low</li> <li>○ Low</li> <li>○ Moderate</li> <li>○ High</li> <li>○ No included studies</li> </ul>		
VALUES	<b>Is there important uncertainty about or variability in how much people value the main outcomes?</b>		

	<ul style="list-style-type: none"> <li>◦ Important uncertainty or variability</li> <li>◦ Possibly important uncertainty or variability</li> <li>◦ Probably no important uncertainty or variability</li> <li>◦ No important uncertainty or variability</li> </ul>		
BALANCE OF EFFECTS	<p><b>Does the balance between desirable and undesirable effects favor the intervention or the comparison?</b></p> <ul style="list-style-type: none"> <li>◦ Favors the comparison</li> <li>◦ Probably favors the comparison</li> <li>◦ Does not favor either the intervention or the comparison</li> <li>◦ Probably favors the intervention</li> <li>◦ Favors the intervention</li> <li>◦ Varies</li> <li>◦ Don't know</li> </ul>		
RESOURCES REQUIRED	<p><b>How large are the resource requirements (costs)?</b></p> <ul style="list-style-type: none"> <li>◦ Large costs</li> <li>◦ Moderate costs</li> <li>◦ Negligible costs and savings</li> <li>◦ Moderate savings</li> <li>◦ Large savings</li> <li>◦ Varies</li> <li>◦ Don't know</li> </ul>		
CERTAINTY OF EVIDENCE OF REQUIRED RESOURCES	<p><b>What is the certainty of the evidence of resource requirements (costs)?</b></p> <ul style="list-style-type: none"> <li>◦ Very low</li> <li>◦ Low</li> <li>◦ Moderate</li> <li>◦ High</li> <li>◦ No included studies</li> </ul>		

COST EFFECTIVENESS	<p><b>Does the cost-effectiveness of the intervention favor the intervention or the comparison?</b></p> <ul style="list-style-type: none"> <li>◦ Favors the comparison</li> <li>◦ Probably favors the comparison</li> <li>◦ Does not favor either the intervention or the comparison</li> <li>◦ Probably favors the intervention</li> <li>◦ Favors the intervention</li> <li>◦ Varies</li> <li>◦ No included studies</li> </ul>		
EQUITY	<p><b>What would be the impact on health equity?</b></p> <ul style="list-style-type: none"> <li>◦ Reduced</li> <li>◦ Probably reduced</li> <li>◦ Probably no impact</li> <li>◦ Probably increased</li> <li>◦ Increased</li> <li>◦ Varies</li> <li>◦ Don't know</li> </ul>		
ACCEPTABILITY	<p><b>Is the intervention acceptable to key stakeholders?</b></p> <ul style="list-style-type: none"> <li>◦ No</li> <li>◦ Probably no</li> <li>◦ Probably yes</li> <li>◦ Yes</li> <li>◦ Varies</li> <li>◦ Don't know</li> </ul>		
FEASIBILITY	<p><b>Is the intervention feasible to implement?</b></p> <ul style="list-style-type: none"> <li>◦ No</li> <li>◦ Probably no</li> <li>◦ Probably yes</li> <li>◦ Yes</li> <li>◦ Varies</li> <li>◦ Don't know</li> </ul>		

## Summary of judgements

	JUDGEMENT							IMPLICATIONS
PROBLEM								
DESIRABLE EFFECTS								
UNDESIRABLE EFFECTS								
CERTAINTY OF EVIDENCE								
VALUES								
BALANCE OF EFFECTS								
RESOURCES REQUIRED								
CERTAINTY OF EVIDENCE OF REQUIRED RESOURCES								
COST EFFECTIVENESS								
EQUITY								
ACCEPTABILITY								
FEASIBILITY								

## Appendix



### **Informed Consent to Participate in Research**

Information to Consider Before Taking Part in this Research Study

**Pro # 00027571**

Researchers at the University of South Florida (USF) study many topics. To do this, we need the help of people who agree to take part in a research study. This form tells you about this research study. We are asking you to take part in a research study that is called: Evaluation of the Group Decision-making Process of Clinical Guidelines Panels. The person who is in charge of this research study is Dr. Benjamin Djulbegovic. This person is called the Principal Investigator.

This study is sponsored by: Agency for Healthcare Research (AHRQ)

### **Purpose of the Study**

GRADE (Grading of Recommendations Assessment, Development and Evaluation) is accepted by more than 100 professional organizations that generate clinical practice guidelines. However, development of guidelines ultimately relies on the group judgment of the panel. Despite the importance of group judgments for issuing guidelines, little work has been done to analyze how this process takes place.

The purpose of this study is to assess how group judgment process reflects the relationship between GRADE and other contextual factors including individual panel member expertise, decision-making styles, etc related to the direction and the strength of recommendations of the guidelines.

### **Why are you being asked to take part?**

We are asking you to take part in this research study because you are a member of a panel for developing clinical guidelines.

### **Study Procedures**

If you take part in this study, you will be asked to complete a couple of surveys in addition to your participation on the guideline development panel. By participating in your Clinical Practice Guidelines panel, you have already agreed to deliberate and issue your recommendations regardless of the proposed study. For this study we ask you to help us formally analyze this process – by completion of the series of questionnaires- prior and during/post meeting- and consenting to recording the panel session discussion.

At the beginning of the survey you will be asked to provide some brief demographic information (your area of specialty, years in practice, your age and your gender, etc). This will be followed by survey related to your decision-making styles. This is estimated to take about 15 minutes.

The next part of the study will relate to your judgments regarding the formulation of guidelines recommendation as per GRADE process. You will be asked to evaluate the presented evidence and make your recommendation for or against the use of the intervention. This is expected to last about 15-20 minutes.

Following the survey, you will meet with your guideline panel, where you will deliberate with other panel members as instructed by your Chair. As per CPG development process, you will be asked to cast your vote either at the end of the meeting, or one week post meeting (depending on your panel). One week after the meeting, we will send you a short follow-up survey (which is expected to take less than 5 minutes of your time) asking you for your overall impression of the guidelines development process.

We will record discussion that occurs during the meeting, which we will then subject to qualitative analysis to identify any new themes of importance for guidelines development that may have not been previously included in the GRADE system. Although by agreeing to participate in the panel, you expressed your willingness to voice your opinion to help improve CPG process, we will deidentify all data prior to analysis. Therefore, there is no way that any particular opinion will be linked back to any individual participant.

Please note that once you complete all answers, the random code will be generated related to your participation and all your identifying information will be erased. Hence, no one will be able to link your answers to you. That is, the surveys will remain anonymous.

### **Alternatives / Voluntary Participation / Withdrawal**

You have the alternative to choose not to participate in this research study.

You should only take part in this study if you want to volunteer; you are free to participate in this research or withdraw at any time. There will be no penalty or loss of benefits you are entitled to receive if you stop taking part in this study. Your decision to not partake in this study will not affect your participation as a member of the CPG panel.

### **Benefits and Risks**

You will receive no benefit from this study.

This research is considered to be minimal risk.

### **Compensation**

We will not pay you for the time you volunteer while being in this study.

### **Privacy and Confidentiality**

We must keep your study records as confidential as possible. It is possible, although unlikely, that unauthorized individuals could gain access to your responses because you are responding online.

Certain people may need to see your study records. By law, anyone who looks at your records must keep them completely confidential. The only people who will be allowed to see these records are: the Principal Investigator and research team data analyst, The University of South Florida Institutional Review Board (IRB) and government offices such as, The Department of Health and Human Services (DHHS).

- It is possible, although unlikely, that unauthorized individuals could gain access to your responses. Confidentiality will be maintained to the degree permitted by the technology used. No guarantees can be made regarding the interception of data sent via the Internet. However, your participation in this online survey involves risks similar to a person's everyday use of the Internet. If you complete and submit an anonymous survey and later request your data be withdrawn, this may or may not be possible as the researcher may be unable to extract anonymous data from the database.

## **Contact Information**

If you have any questions about your rights as a research participant, please contact the USF IRB at (813) 974-5638 or contact by email at [RSCH-IRB@usf.edu](mailto:RSCH-IRB@usf.edu). If you have questions regarding the research, please contact the Principal Investigator at Dr. Benjamin Djulbegovic at USF Health, phone: (813) 396-2349, email: [bdjulbeg@health.usf.edu](mailto:bdjulbeg@health.usf.edu).

We may publish what we learn from this study. If we do, we will not let anyone know your name. We will not publish anything else that would let people know who you are. You can print a copy of this consent form for your records.

I freely give my consent to take part in this study. I understand that by proceeding with this survey that I am agreeing to take part in research and I am 18 years of age or older.

[ <https://usf.qualtrics.com> ]

**A) Part I: Baseline survey to be administered to all participants before the guideline panel meets**

Thank you for choosing to participate in our study. Below is a baseline survey that includes several demographic questions, a set of standard scales that have been validated in prior studies used to evaluate decision making style<sup>1-8</sup>, a GRADE evaluation tool, and your recommendations. The survey is expected to take approximately 15 to 20 minutes to complete.

**Demographics**

What is your primary role on the current panel *(please select only one)*

---

Prior to this panel, have you previously participated in the development of guidelines?

☐ Yes ☐ No

If you answered 'yes', how many guidelines have you participated in:

In what capacity did you participate in previous guidelines? *(please select all that apply)*

☐ Panel member

☐ Methodologist

☐ Panel Chair

☐ Other:

---

What is your formal education (e.g. MD, RN, MSc, MPH, PhD, etc.)?

---

Do you have formal training in health research methodology/epidemiology/biostatistics?

☐ Never completed formal training

☐ Completed some formal training but do not have graduate degree

☐ Earned MSc degree

☐ Earned PhD degree

---

Indicate your field of work *(please select only one)*

---

Please specify 'other' areas of expertise:



---

How many years of experience do you have in your field?

---

If you are a clinical expert, compared to other people you know in your field, how would you rate your level of expertise regarding the recommendations you are most knowledgeable about?

- ☐ Higher than others
- ☐ About the same as others
- ☐ Lower than others

---

If you are a clinical expert, how many patients do you see per month that match the population affected by the guideline you have the most expertise in?

- ☐ None
- ☐ 1 to 5
- ☐ 6 to 10
- ☐ 11 to 15
- ☐ More than 15

---

What is your age?

---

What is your gender?

- ☐ Male
- ☐ Female

---

Do you have any financial conflict of interest with respect to the guideline recommendations?

☐ Yes ☐ No

If you answered 'yes', please briefly explain below:

---

Do you have any intellectual conflict of interest with respect to the guideline recommendations?

☐ Yes ☐ No

If you answered 'yes', please briefly explain below:

---

Do you have any institutional conflict of interest with respect to the guideline recommendations?

☐ Yes ☐ No

If you answered 'yes', please briefly explain below:

---

Do you believe that these guidelines have particular social implications which may affect one or more vulnerable populations [e.g. women, children, racial and ethnic minorities, populations with special health care needs (chronic illness, disabilities, and end of life), the elderly, low-income, inner-city, and rural populations]?

☐ Yes ☐ No

If you answered 'yes', please briefly explain below:

---

Do you feel that you are expected to conform or inappropriately pressured to vote (issue recommendation) in a particular way?

☐ Yes

☐ No

If you answered 'yes', please indicate where the pressure is coming from:

☐ Peers

☐ Politicians

☐ Regulators

☐ Government

☐ Insurance

☐ Society at large

☐ Other, please briefly explain:

## Assessment of individual differences (or, traits) in decision-making

All scales and items within scales will be randomly presented to control for order effects.

**Objectivism Scale.** <sup>4</sup> (*The scale is identified here for the reviewers' convenience. This label will not be presented to participants.*)

Below are several statements that describe how various people make decisions in general. Read each statement carefully and think about the extent to which the statement describes you. Use the following rating scale to indicate your responses.

0	1	2	3	4
Not at all characteristic of me	Slightly characteristic of me	Moderately characteristic of me	Very characteristic of me	Extremely characteristic of me

1. \_\_\_\_ I seek as much information as possible before making decisions.
2. \_\_\_\_ I think the answers to most questions in life can be found through careful, objective analysis of the situation.
3. \_\_\_\_ I do not like to be too objective in the way I look at things.
4. \_\_\_\_ Trying to be highly objective and rational does not improve my ability to make good decisions.
5. \_\_\_\_ I see myself as a rational and objective person.
6. \_\_\_\_ After I make a decision, it is often difficult for me to give logical reasons for it.
7. \_\_\_\_ I gather as much information as possible before making decisions.
8. \_\_\_\_ The solution to many problems in life can not be found through an intellectual examination of the facts.
9. \_\_\_\_ I try to employ a cool-headed, objective approach when making decisions about my life.
10. \_\_\_\_ I am only confident of decisions that are made after careful analysis of all available information.
11. \_\_\_\_ I tend not to be particularly objective or logical in my approach to life.

**Rational-Experiential Inventory.** <sup>5</sup> (*The scale is identified here for the reviewers' convenience. This label will not be presented to participants.*)

Below are several statements that describe how various people make decisions in general. Read each statement carefully and think about the extent to which the statement describes you. Use the following rating scale to indicate your responses.

0	1	2	3	4
Not at all characteristic of me	Slightly characteristic of me	Moderately characteristic of me	Very characteristic of me	Extremely characteristic of me

1. \_\_\_\_\_ I would not want to depend on anyone who described himself or herself as intuitive.
2. \_\_\_\_\_ My snap judgments are probably not as good as most people's.
3. \_\_\_\_\_ I tend to use my heart as a guide for my actions.
4. \_\_\_\_\_ I can usually feel when a person is right or wrong, even if I can't explain how I know.
5. \_\_\_\_\_ I suspect my hunches are inaccurate as often as they are accurate.
6. \_\_\_\_\_ I try to avoid situations that require thinking in depth about something.
7. \_\_\_\_\_ I'm not that good at figuring out complicated problems.
8. \_\_\_\_\_ When it comes to trusting people, I can usually rely on my gut feelings.
9. \_\_\_\_\_ I enjoy intellectual challenges.
10. \_\_\_\_\_ I am not very good at solving problems that require careful logical analysis.
11. \_\_\_\_\_ I don't like to have to do a lot of thinking.
12. \_\_\_\_\_ I often go by my instincts when deciding on a course of action.
13. \_\_\_\_\_ I trust my initial feelings about people.
14. \_\_\_\_\_ If I were to rely on my gut feelings, I would often make mistakes.
15. \_\_\_\_\_ I don't like situations in which I have to rely on intuition.
16. \_\_\_\_\_ Knowing the answer without having to understand the reasoning behind it is good enough for me.
17. \_\_\_\_\_ I don't reason well under pressure.
18. \_\_\_\_\_ I am much better at figuring things out logically than most people.
19. \_\_\_\_\_ I have a logical mind.
20. \_\_\_\_\_ I enjoy thinking in abstract terms.
21. \_\_\_\_\_ Thinking hard and for a long time about something gives me little satisfaction.
22. \_\_\_\_\_ I think there are times when one should rely on one's intuition.
23. \_\_\_\_\_ I think it is foolish to make important decisions based on feelings.

**Rational-Experiential Inventory (continued).** (*The scale is identified here for the reviewers' convenience. This label will not be presented to participants.*)

0	1	2	3	4
Not at all characteristic of me	Lightly characteristic of me	Moderately characteristic of me	Very characteristic of me	Extremely characteristic of me

24. \_\_\_\_\_ I don't think it is a good idea to rely on one's intuition for important decisions.
25. \_\_\_\_\_ I generally don't depend on my feelings to help me make decisions.
26. \_\_\_\_\_ I hardly ever go wrong when I listen to my deepest gut feelings to find an answer.
27. \_\_\_\_\_ I have no problem thinking things through carefully.
28. \_\_\_\_\_ Using logic usually works well for me in figuring out problems in my life.
29. \_\_\_\_\_ I usually have clear, explainable reasons for my decisions.
30. \_\_\_\_\_ Learning new ways to think would be very appealing to me.
31. \_\_\_\_\_ I like to rely on my intuitive impressions.
32. \_\_\_\_\_ I don't have a very good sense of intuition.
33. \_\_\_\_\_ Using my gut feelings usually works well for me in figuring out problems in my life.
34. \_\_\_\_\_ I believe in trusting my hunches.
35. \_\_\_\_\_ Intuition can be a very useful way to solve problems.
36. \_\_\_\_\_ I enjoy solving problems that require hard thinking.
37. \_\_\_\_\_ Thinking is not my idea of an enjoyable activity.
38. \_\_\_\_\_ I am not a very analytical thinker.
39. \_\_\_\_\_ Reasoning things out carefully is not one of my strong points.
40. \_\_\_\_\_ I prefer complex problems to simple problems.

**Intolerance for Ambiguity Scale.** <sup>2</sup> (*The scale is identified here for the reviewers' convenience. This label will not be presented to participants.*)

Below are several general statements regarding how people perceive and think about the world around them. Please read each statement carefully. Indicate the extent to which you agree or disagree with each statement using the following rating scale:

1	2	3	4	5	6
Strongly disagree	Disagree	Slightly disagree	Slightly agree	Agree	Strongly agree

1. \_\_\_\_\_ An expert who doesn't come up with a definite answer probably doesn't know very much.
2. \_\_\_\_\_ Teachers or supervisors who hand out vague assignments give a chance for one to show initiative and originality.
3. \_\_\_\_\_ People who fit their lives to a schedule probably miss most of the joy of living.
4. \_\_\_\_\_ Often the most interesting and stimulating people are those who don't mind being different and original.
5. \_\_\_\_\_ It is more fun to tackle a complicated problem than to solve a simple one.
6. \_\_\_\_\_ In the long run it is possible to get more done by tackling small, simple problems rather than large and complicated ones.
7. \_\_\_\_\_ A good job is one where what is to be done and how it is to be done are always clear.
8. \_\_\_\_\_ A person who leads an even, regular life in which few surprises or unexpected happenings arise, really has a lot to be grateful for.
9. \_\_\_\_\_ What we are used to is always preferable to what is unfamiliar to us.
10. \_\_\_\_\_ People who insist upon a "yes" or "no" answer just don't know how complicated things really are.
11. \_\_\_\_\_ There is really no such thing as a problem that can't be solved.
12. \_\_\_\_\_ Many of our most important decisions are based upon insufficient information.
13. \_\_\_\_\_ I like parties where I know most of the people more than ones where all or most of the people are complete strangers.
14. \_\_\_\_\_ I would like to live in a foreign country for a while.
15. \_\_\_\_\_ The sooner we all acquire similar values and ideals the better.
16. \_\_\_\_\_ A good teacher is one who makes you wonder about your own way of looking at things.

**Decision Making Tendency Inventory<sup>9</sup>** *(The scale is identified here for the reviewers' convenience. This label will not be presented to participants.)*

Below are several statements that describe how various people make decisions in general. Read each statement carefully and think about the extent to which the statement describes you. Use the following rating scale to indicate your responses.

0	1	2	3	4
Not at all characteristic of me	Slightly characteristic of me	Moderately characteristic of me	Very characteristic of me	Extremely characteristic of me

**Satisficing**

1. \_\_\_\_\_ In every area, I try to achieve results that are satisfactory for me
2. \_\_\_\_\_ In studying or working, I tend to choose solutions that guarantee satisfactory results for me
3. \_\_\_\_\_ When I make decisions, I spend the time required to choose an alternative that is satisfactory for me
4. \_\_\_\_\_ In studying or working, I spend the time required to choose solutions that meet my needs
5. \_\_\_\_\_ If I am happy with my work, I do not seek better opportunities
6. \_\_\_\_\_ In choosing between alternatives, I stop at the first that works for me
7. \_\_\_\_\_ I do not ask for more than what satisfies me
8. \_\_\_\_\_ When I watch TV or listen to the radio, I tend to follow the first program that I find interesting



**Maximizing Tendency Inventory<sup>10</sup>** *(The scale is identified here for the reviewers' convenience. This label will not be presented to participants.)*

Below are several statements that describe how various people make decisions in general. Read each statement carefully and think about the extent to which the statement describes you. Use the following rating scale to indicate your responses.

0	1	2	3	4
Not at all characteristic of me	Slightly characteristic of me	Moderately characteristic of me	Very characteristic of me	Extremely characteristic of me

1. \_\_\_\_\_ No matter what I do, I have the highest standards for myself.
2. \_\_\_\_\_ I never settle for second best.
3. \_\_\_\_\_ No matter what it takes, I always try to choose the best thing.
4. \_\_\_\_\_ I don't like having to settle for "good enough."
5. \_\_\_\_\_ I am a maximizer.
6. \_\_\_\_\_ I will wait for the best option, no matter how long it takes.
7. \_\_\_\_\_ I never settle.

**Maximization Inventory<sup>11</sup>** *(The scale is identified here for the reviewers' convenience. This label will not be presented to participants.)*

Below are several statements describing how people think and feel about decision making. Please read each statement carefully. Indicate the extent to which you agree or disagree with each statement using the following rating scale:

1	2	3	4	5	6
Strongly disagree	Disagree	Slightly disagree	Slightly agree	Agree	Strongly agree

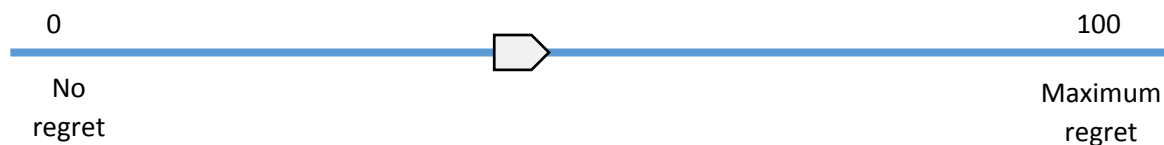
1. \_\_\_\_ I usually try to find a couple of good options and then choose between them.
2. \_\_\_\_ I usually have a hard time making even simple decisions.
3. \_\_\_\_ I can't come to a decision unless I have carefully considered all of my options.
4. \_\_\_\_ At some point you need to make a decision about things.
5. \_\_\_\_ I am usually worried about making a wrong decision.
6. \_\_\_\_ I take time to read the whole menu when dining out.
7. \_\_\_\_ In life I try to make the most of whatever path I take.
8. \_\_\_\_ I often wonder why decisions can't be more easy.
9. \_\_\_\_ I will continue shopping for an item until it reaches all of my criteria.
10. \_\_\_\_ There are usually several good options in a decision situation.
11. \_\_\_\_ I often put off making a difficult decision until a deadline.
12. \_\_\_\_ I usually continue to search for an item until it reaches my expectations.
13. \_\_\_\_ I try to gain plenty of information before I make a decision, but then I go ahead and make it.
14. \_\_\_\_ I often experience buyer's remorse.
15. \_\_\_\_ When shopping, I plan on spending a lot of time looking for something.
16. \_\_\_\_ Good things can happen even when things don't go right at first.
17. \_\_\_\_ I often think about changing my mind after I have already made my decision.
18. \_\_\_\_ When shopping, if I can't find exactly what I'm looking for, I will continue to search for it.

19. \_\_\_\_ I can't possibly know everything before making a decision.
20. \_\_\_\_ The hardest part of making a decision is knowing I will have to leave the item I didn't choose behind.
21. \_\_\_\_ I find myself going to many different stores before finding the thing I want.
22. \_\_\_\_ I do not agonize over decisions.
23. \_\_\_\_ I just won't make a decision until I am comfortable with the process.
24. \_\_\_\_ All decisions have pros and cons.
25. \_\_\_\_ I often change my mind several times before making a decision.
26. \_\_\_\_ When shopping for something, I don't mind spending several hours looking for it.
27. \_\_\_\_ I know that if I make a mistake in a decision that I can go "back to the drawing board."
28. \_\_\_\_ It's hard for me to choose between two good alternatives.
29. \_\_\_\_ I take the time to consider all alternatives before making a decision.
30. \_\_\_\_ I accept that life often has uncertainty.
31. \_\_\_\_ Sometimes I procrastinate in deciding even if I have a good idea of what decision I will make.
32. \_\_\_\_ When I see something that I want, I always try to find the best deal before purchasing it.
33. \_\_\_\_ I find myself often faced with difficult decisions.
34. \_\_\_\_ If a store doesn't have exactly what I'm shopping for, then I will go somewhere else.

### Anticipated Regret Scale <sup>3,6,7</sup>

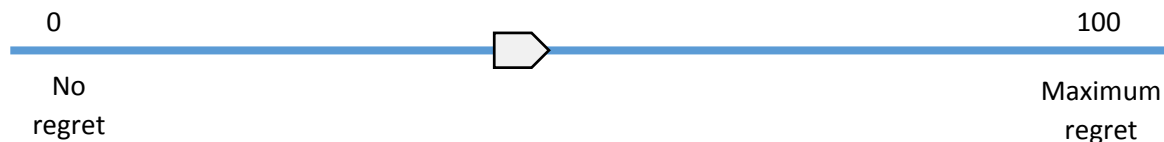
Read each statement below carefully and think about the extent to which the statement describes your attitude. Use the scale below to indicate your responses or enter a number from 0 (no regret) to 100 (maximum regret).

When you cast your "vote" as a **STRONG** recommendation **FOR** a health intervention how much regret would you feel if it turned out to be unnecessary and possibly harmful **(which will lead to more undesirable than desirable consequences)**?



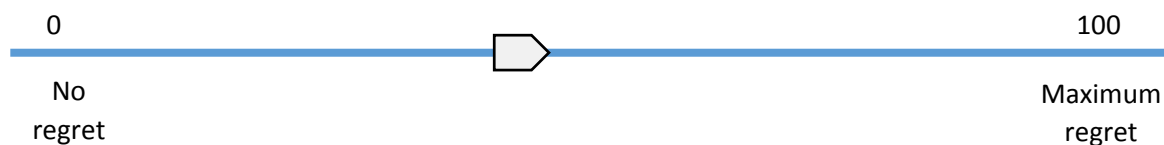
Enter your level of regret:

When you cast your "vote" as a **STRONG** recommendation **AGAINST** a health intervention how much regret would you feel if you failed to recommend a health intervention that could improve patient outcomes **(which will fail to lead to more desirable than undesirable consequences)**?



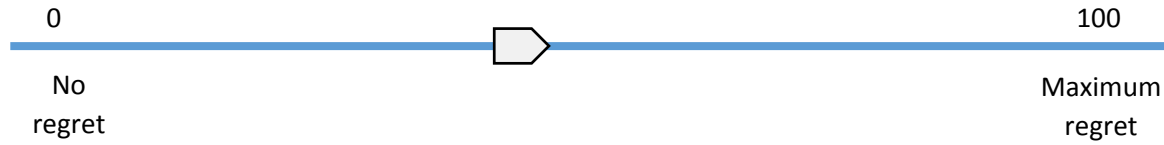
Enter your level of regret:

When you cast your "vote" as a **WEAK** recommendation **FOR** a health intervention how much regret would you feel if it turn out to be unnecessary and possibly harmful **(which will lead to more undesirable than desirable consequences)**?



Enter your level of regret:

When you cast your "vote" as the WEAK recommendation AGAINST a health intervention how much regret would you feel if you failed to recommend a health intervention that could improve patient outcomes **(which will fail to lead to more desirable than undesirable consequences)**?



Enter your level of regret:

## A) Part II: GRADE data presentation (BEFORE the panel meeting)

You are about to be asked to make a recommendation for or against the use of intervention \_\_\_\_\_ for condition \_\_\_\_\_. Before making your recommendation please review the summary assessment of key GRADE domains already provided for you. If you disagree with any domain summary, please indicate your opinion below each statement.

**1) Is the problem a priority?**

- ☐ No
- ☐ Probably no
- ☐ Probably yes
- ☐ Yes
- ☐ Varies
- ☐ Don't know

If you disagree with the assessment above, please indicate your perceived assessment of problem priority:

**2) How substantial are the desirable anticipated effects?**

- ☐ Trivial
- ☐ Small
- ☐ Moderate
- ☐ Large
- ☐ Varies
- ☐ Don't know

If you disagree with the assessment above, please indicate your perceived assessment of desirable anticipated effects:

**3) How substantial are the undesirable anticipated effects?**

- ☐ Large
- ☐ Moderate
- ☐ Small
- ☐ Trivial
- ☐ Varies
- ☐ Don't know

If you disagree with the assessment above, please indicate your perceived assessment of undesirable anticipated effects:

- ☐ Very low

**4) What is the overall certainty (quality) of the evidence of effects?**

- ☐ Low
- ☐ Moderate
- ☐ High
- ☐ Varies
- ☐ Don't know

If you disagree with the assessment above, please indicate your perceived assessment of certainty (quality) of the evidence of effects:

**5) Is there important uncertainty about or variability in how much people value the main outcomes?**

- ☐ Important uncertainty or variability
- ☐ Possibly important uncertainty or variability
- ☐ Probably no important uncertainty or variability
- ☐ No important uncertainty or variability

If you disagree with the assessment above, please indicate your perceived assessment of how much people value the main outcomes:

**6) Does the balance between desirable and undesirable effects favor the intervention or the comparison?**

- ☐ Favors the comparison
- ☐ Probably favors the comparison
- ☐ Does not favor either the intervention or the comparison
- ☐ Probably favors the intervention
- ☐ Favors the intervention
- ☐ Varies
- ☐ Don't know

If you disagree with the assessment above, please indicate your perceived assessment of balance between desirable and undesirable effects:

**7) How large are the resource requirements (costs)?**

- ☐ Large costs
- ☐ Moderate costs
- ☐ Negligible costs and savings
- ☐ Moderate savings

- ☐ Large savings
- ☐ Varies
- ☐ Don't know

If you disagree with the assessment above, please indicate your perceived assessment of the resource requirements (costs):

**8) What would be the impact on health equity?**

- ☐ Reduced
- ☐ Probably reduced
- ☐ Probably no impact
- ☐ Probably increased
- ☐ Increased
- ☐ Varies
- ☐ Don't know

If you disagree with the assessment above, please indicate your perceived assessment of the impact on health equity:

**9) Is the intervention acceptable to key stakeholders?**

- ☐ No
- ☐ Probably no
- ☐ Probably yes
- ☐ Yes
- ☐ Varies
- ☐ Don't know

If you disagree with the assessment above, please indicate your perceived assessment of the intervention's acceptability to stakeholders:



**10) Is the intervention  
feasible to implement?**

- ☐ No
- ☐ Probably no
- ☐ Probably yes
- ☐ Yes
- ☐ Varies
- ☐ Don't know

If you disagree with the assessment above, please indicate your perceived assessment of feasibility of the intervention's implementation:

### Recommendation vote

Please indicate your recommendation related to the use of intervention \_\_\_\_\_ for condition \_\_\_\_\_.

- ☐ I STRONGLY recommend FOR using intervention \_\_\_\_\_ for condition \_\_\_\_\_.
  - ☐ I WEAKLY recommend FOR using intervention \_\_\_\_\_ for condition \_\_\_\_\_.
  - ☐ I WEAKLY recommend AGAINST using intervention \_\_\_\_\_ for condition \_\_\_\_\_.
  - ☐ I STRONGLY recommend AGAINST intervention \_\_\_\_\_ for condition \_\_\_\_\_.
- 

B) Participants of guideline **panels meet** to discuss recommendations. The deliberations are recorded. [Details related to specific guideline recommendations will also be recorded in the PICO (patients, interventions, comparators, outcomes) format]. <sup>1</sup>Immediately **during, or within a week following the guideline panel meeting**, the participants are asked to complete the following survey.

---

<sup>1</sup> This will be panel dependent; some panels will use GDP software, so all “votes” will be recorded **during** the meeting.

## GRADE data presentation (DURING/AFTER the panel meeting)

You are about to be asked to make a recommendation for or against the use of intervention \_\_\_\_\_ for condition \_\_\_\_\_. Before making your recommendation please review the summary assessment of key GRADE domains already provided for you. If you disagree with any domain summary, please indicate your opinion below each statement.

**1) Is the problem a priority?**

- ☐ No
- ☐ Probably no
- ☐ Probably yes
- ☐ Yes
- ☐ Varies
- ☐ Don't know

If you disagree with the assessment above, please indicate your perceived assessment of problem

priority:

**2) How substantial are the desirable anticipated effects?**

- ☐ Trivial
- ☐ Small
- ☐ Moderate
- ☐ Large
- ☐ Varies
- ☐ Don't know

If you disagree with the assessment above, please indicate your perceived assessment of desirable anticipated effects:

**3) How substantial are the undesirable anticipated effects?**

- ☐ Large
- ☐ Moderate
- ☐ Small
- ☐ Trivial
- ☐ Varies
- ☐ Don't know

If you disagree with the assessment above, please indicate your perceived assessment of undesirable anticipated effects:

- ☐ Very low

**4) What is the overall certainty (quality) of the evidence of effects?**

- ☐ Low
- ☐ Moderate
- ☐ High
- ☐ Varies
- ☐ Don't know

If you disagree with the assessment above, please indicate your perceived assessment of certainty (quality) of the evidence of effects:

**5) Is there important uncertainty about or variability in how much people value the main outcomes?**

- ☐ Important uncertainty or variability
- ☐ Possibly important uncertainty or variability
- ☐ Probably no important uncertainty or variability
- ☐ No important uncertainty or variability

If you disagree with the assessment above, please indicate your perceived assessment of how much people value the main outcomes:

**6) Does the balance between desirable and undesirable effects favor the intervention or the comparison?**

- ☐ Favors the comparison
- ☐ Probably favors the comparison
- ☐ Does not favor either the intervention or the comparison
- ☐ Probably favors the intervention
- ☐ Favors the intervention
- ☐ Varies
- ☐ Don't know

If you disagree with the assessment above, please indicate your perceived assessment of balance between desirable and undesirable effects:

**7) How large are the resource requirements (costs)?**

- ☐ Large costs
- ☐ Moderate costs
- ☐ Negligible costs and savings
- ☐ Moderate savings

- ☐ Large savings
- ☐ Varies
- ☐ Don't know

If you disagree with the assessment above, please indicate your perceived assessment of the resource requirements (costs):

**8) What would be the impact on health equity?**

- ☐ Reduced
- ☐ Probably reduced
- ☐ Probably no impact
- ☐ Probably increased
- ☐ Increased
- ☐ Varies
- ☐ Don't know

If you disagree with the assessment above, please indicate your perceived assessment of the impact on health equity:

**9) Is the intervention acceptable to key stakeholders?**

- ☐ No
- ☐ Probably no
- ☐ Probably yes
- ☐ Yes
- ☐ Varies
- ☐ Don't know

If you disagree with the assessment above, please indicate your perceived assessment of the intervention's acceptability to stakeholders:

- ☐ No

**10) Is the intervention  
feasible to implement?**

- ☐ Probably no
- ☐ Probably yes
- ☐ Yes
- ☐ Varies
- ☐ Don't know

If you disagree with the assessment above, please indicate your perceived assessment of feasibility of the intervention's implementation:

### Recommendation vote

Please indicate your recommendation related to the use of intervention \_\_\_\_\_ for condition \_\_\_\_\_.

- ☐ I STRONGLY recommend FOR using intervention \_\_\_\_\_ for condition \_\_\_\_\_.
- ☐ I WEAKLY recommend FOR using intervention \_\_\_\_\_ for condition \_\_\_\_\_.
- ☐ I WEAKLY recommend AGAINST using intervention \_\_\_\_\_ for condition \_\_\_\_\_.
- ☐ I STRONGLY recommend AGAINST intervention \_\_\_\_\_ for condition \_\_\_\_\_.

### C) Follow-up survey to be administered to all participants within 1 week of guideline panel meeting

Thank you for participating in our study. Below is a brief follow-up related to your participation in development of clinical practice guidelines and an assessment of how you feel about your decision. The survey is expected to take less than 5 minutes to complete but you can take as long as you wish to answer questions the asked in the survey.

#### Post Decisional Regret Scale

##### 1. Brehaut scale <sup>1</sup>

Please reflect on the final vote that your panel made for **all recommendations** for or against a health intervention. Please indicate how strongly you agree or disagree with the statements below with respect to majority (>80%) of your recommendations by using the following rating scale:

1	2	3	4	5
Strongly agree	Agree	Neither agree nor disagree	Disagree	Strongly disagree

1. \_\_\_\_ It was the right decision (recommendation).
2. \_\_\_\_ I regret the choices (recommendations) that were made.
3. \_\_\_\_ I would go (vote) for the same choices (recommendations) if I had to do it over again.
4. \_\_\_\_ The choices (recommendations) will do a lot of harm.
5. \_\_\_\_ The decisions (recommendations) were wise.

If you believe that you should have made some recommendations differently, please briefly explain below:

##### 2. Do you feel that you were expected to conform or were inappropriately pressured to vote (issue recommendation) in a particular way?

- ☐ Yes ☐ No

If you answered 'yes', please indicate where the pressure came from:

- ☐ Peers
- ☐ Politicians
- ☐ Regulators
- ☐ Government
- ☐ Insurance
- ☐ Society at large

☐ Other, please briefly explain:





**3. Did you feel that discussion/beliefs of a single individual had disproportionate influence on the guidelines development process?**

☐ **Yes**

☐ **No**

If you answered 'yes', please briefly explain below:

**4. Your overall comments**

Please provide us with any other thoughts you have related to the guidelines recommendations process in which you participated. We would appreciate receiving both "positive" and "negative" comments.

---

**D) Closing questions to be completed by the guideline chair after peer review process is completed**

**1. Have guideline recommendations changed as a result of the peer-review process?**

☐ **Yes**

☐ **No**

If you answered 'yes', please indicate how many recommendations changed and comment on what changes occurred (e.g. the quality of evidence was reassessed, the strength of recommendation was modified, etc.)

## References

1. Brehaut JC, O'Connor AM, Wood TJ, et al. Validation of a decision regret scale. *Med Decis Making*. Jul-Aug 2003;23(4):281-292.
2. Budner S. Intolerance of ambiguity as a personality variable. *Journal of Personality*. 1962;30:29-50.
3. Connolly TaR, J. *Regret in Cancer-Related Decisions*. 2005.
4. Leary MR, Shepperd JA, McNeil MS, Jenkins TB, Barnes BD. Objectivism in information utilization: theory and measurement. *Journal of Personality Assessment*. 1986;50(1):32-43.
5. Pacini R, Epstein S. The relation of rational and experiential information processing styles to personality, basic beliefs, and the ratio-bias phenomenon *Journal of Personality and Social Psychology*. 1999;76:972-987.
6. Sandberg T, Conner M. A mere measurement effect for anticipated regret: impacts on cervical screening attendance. *Br J Soc Psychol*. Jun 2009;48(Pt 2):221-236.
7. Ziarnowski KL, Brewer NT, Weber B. Present choices, future outcomes: anticipated regret and HPV vaccination. *Prev Med*. May 2009;48(5):411-414.
8. Djulbegovic B, Beckstead JW, Elqayam S, et al. Evaluation of Physicians' Cognitive Styles. *Med Decis Making*. Apr 10 2014;34(5):627-637.
9. Misuraca R, Faraci P, Gangemi A, Carmeci FA, Miceli S. The Decision Making Tendency Inventory: A new measure to assess maximizing, satisficing, and minimizing. *Personality and Individual Differences*. 10// 2015;85:111-116.
10. Dalal DK, Diab DL, Zhu X, Hwang T. Understanding the Construct of Maximizing Tendency: A Theoretical and Empirical Evaluation. *Journal of Behavioral Decision Making*. 2015;28(5):437-450.
11. Turner BM, Rim HB, Betz NE, Nygren TE. The Maximization Inventory *Judgment and Decision Making*. 2012;7(1):48-60.